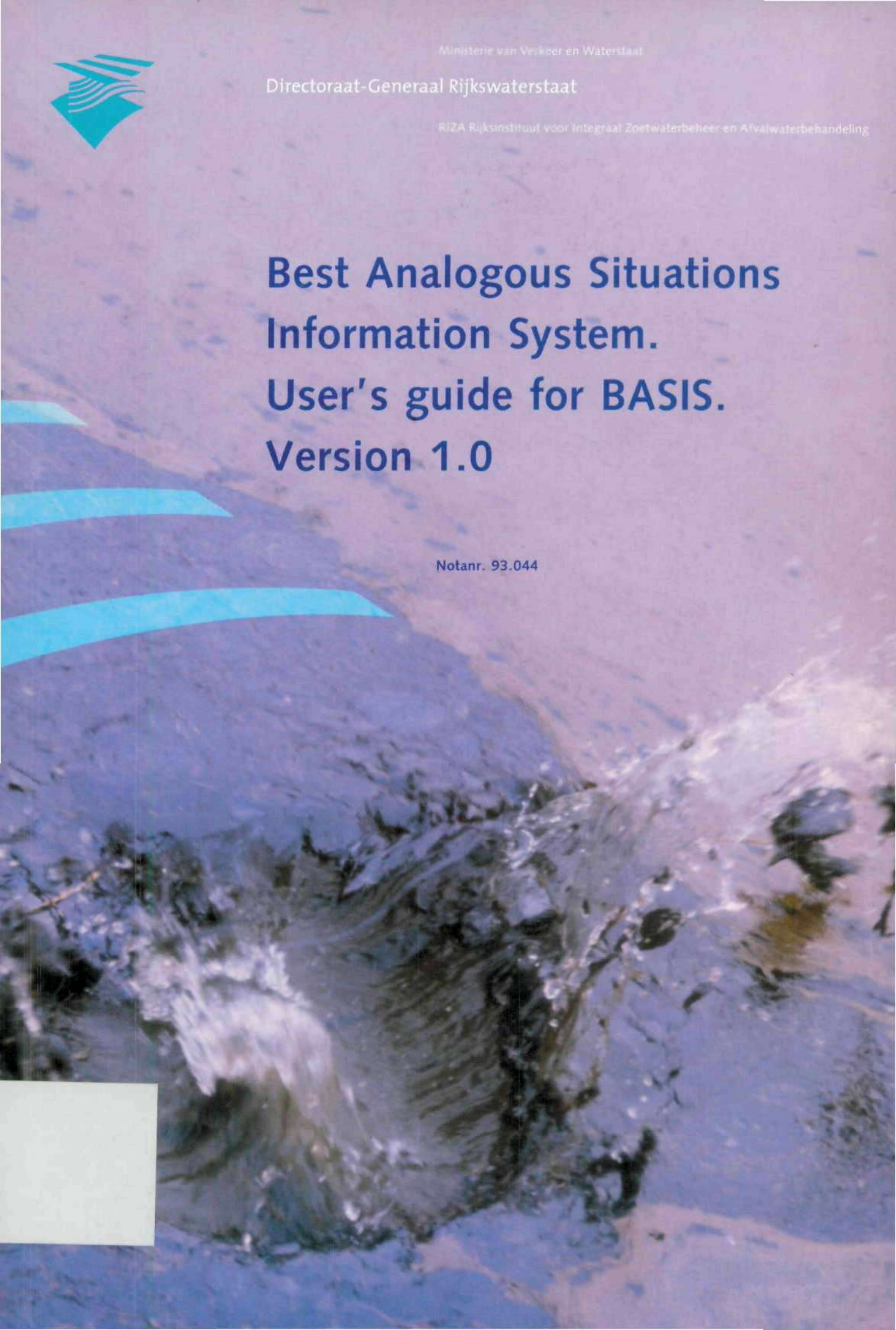
Ministerie van Verkeer en Waterstaat

Directoraat-Generaal Rijkswaterstaat

RIZA Rijksinstituut voor Integraal Zoetwaterbeheer en Afvalwaterbehandeling

Best Analogous Situations Information System. User's guide for BASIS. Version 1.0

Notanr. 93.044



ministerie van verkeer en waterstaat

rijkswaterstaat

riza
rijksinstituut
voor integraal zoetwaterbeheer
en afvalwaterbehandeling
tel. 03200-70411, fax. 03200-49218
doorkiesnummer 70713

Best Analogous Situations Information System. User's guide for BASIS. Version 1.0

NOTA **93.044**

auteur(s) E. H. van Nes & M. Scheffer

datum December 1993

ISBN 90 369 0103 0

CONTENTS

SUMMARY 5

1. INTRODUCTION 7
   1.1. What is BASIS? 7
   1.2. Why WAP? 7

2. INSTALLATION 9
   2.1. System requirements 9
   2.2. Using INSTALL 9
   2.3. Running BASIS 10
   2.4. Getting help 10
   2.5. Special keys and how to use the mouse 11

3. METHODS 13
   3.1. How to find analogies? 13
   3.2. How to predict a response variable? 18
   3.3. How accurate is the prediction? 19
   3.4. Optimization of the prediction 21

4. TUTORIAL: AN EXAMPLE OF A PREDICTION 23
   4.1. Selecting your variables 23
   4.2. The main menu 26
   4.3. Filtering the database 27
   4.4. Finding analogies 30
   4.5. Predicting a response 31
   4.6. Viewing relations between variables 33
   4.7. Calculate percentage explained variance (R^2) 36
   4.8. Optimizing your model 37
   4.9. Saving and loading your model 38

5. GLOSSARY 39

5. REFERENCES 43

APPENDICES 45
   Appendix 1. Start-up file 45
   Appendix 2. How to enter new data in the database? 46
   Appendix 3. Format of the grid files 47
   Appendix 4. Error messages 48
   Appendix 5. The database 50

LIST OF FIGURES

1. How to use vertical scroll bars 12
2. (a) Schedule of the method of predictions (b) Example 14
3. (a) Schedule of the method of cross validation (b) Example 20
4. The input screen 23
5. To get help on any topic press twice [F1] and choose a topic in the help index 24
6. After selection of the variables the screen looks like this 25
7. The main menu. 26
8. A logical expression can be entered to select part of the database. 29
11. Click to examine individual points 33
12. Two dimensional plot with the WAP predictions. 34
13. Three dimensional plot of two conditional variables and one response variable. 35
14. Test predictions and calculate the percentage explained variance (adjusted R^2). 36
15. Optimization of weights. 37

LIST Of TABLES

1. Logical operators are used to combine to relational equations that can be TRUE or FALSE. How these operators work is explained in this table. 28

SUMMARY

BASIS (Best Analogous Situations Information System) is a graphical oriented computer program that runs on Personal Computers. It aims to be an aid for ecologists for making predictions of the response of lakes to measures.

The user of the program enters relevant properties of the lake in question. The program selects in a data set the most similar lakes based on these variables. The database includes physical and chemical parameters, information on morphology, abundances of main groups of algae, zooplankton and fish and coverage with macrophytes. Information of 112 lakes from Holland is included.

The selected similar lakes are visualized in a way that differences and similarities are easily recognized. A prediction of the unknown lake is made by averaging the responses of the most similar lakes. The program includes a test of how accurate these predictions are and an iterative algorithm to optimize the prediction.

Besides the predictions, the program includes an option for exploratory analysis of the data set. Two and three dimensional plots can be made, individual points can easily be examined and selections can be highlighted.

The computer program is menu-driven and includes an extended on-line help system with more than 80 entries. The use of a mouse is fully supported.

This manual describes:

1. How to install BASIS on your PC.
2. How to get started with the program (special keys, the use of a mouse).
3. How precisely the predictions are made and how the predictions can be optimized.
4. How to enter new data in the data-set.
5. The used terminology in a glossary.

In an extended tutorial, you can learn step-by-step what the most important options of the program are.

1. INTRODUCTION

1.1. What is BASIS?

A central theme in water management research is the problem how to select the optimal measure for restoration.

Solving the problem requires prognosis of the expected response of the ecosystem. Up till now, the dominant approach has been to combine all available quantitative information on important processes in computer models for prediction. Such a compilation of information, however, is accompanied by a compilation of errors and uncertainties, which is the main reason that this approach is losing popularity.

Most experts do better than most models. We think that this is because experts tend to reach their conclusion in a safer way than models do.

Experts use analogies: the response of a specific lake to removal of fish is expected to be similar to that of comparable lakes where this has been done before. Obviously all lakes are different, so that some nuance has to be made also.

We formalized the search for analogous lakes. The computer selects the most similar lakes from some relevant variables. The characteristics of the selected cases are visualized in such way that differences and similarities are easily recognized.

This computer information system that we call *Best Analogous Situations Information System (BASIS)* aims to be an aid for ecologists for making predictions of the response of aquatic systems to measures. Besides the WAP technique, the program includes an option for exploratory analysis of the database. Two and three dimensional plots can be made and outliers can easily be examined and selections can be highlighted.

The computer program is menu driven and includes an extended on-line help system with more than 80 entries. The use of a mouse is fully supported.

The computer program is written in Borland Pascal 7.0 (some procedures of Turbo Professional 5.0 are used for the real mode version).

1.2. Why WAP?

WAP (Weighted Analogies Prediction) is a technique to make a prediction based on information of analogous cases.

The technique involves two steps:

1. Find analogous lakes in a database.
2. Predict the response of the question lake using the most analogous lakes.

The advantages of the WAP technique are:

1. No prior information or assumptions about the nature of relations between variables are needed.
2. It is easy to find and browse through the available information of the most analogous lakes.
3. It is the starting point of the LABDA approach (Largely Analogous But Differences Also) (Scheffer, 1991).

   The LABDA approach is a way of predicting the response of a lake.

   It involves two steps:

   A. Rough estimations using analogous cases (Largely Analogous)

   B. Fine tuning of the prediction by quantitative models that predict only the differences of the question lake from the analogous cases (But Differences Also).

2. INSTALLATION

2.1. System requirements

BASIS is a graphical oriented computer program. It runs on any DOS system if one of the following graphical drivers (or compatible drivers) is available:

VGA (or *SVGA*) (recommended)

EGA

Hercules monochrome

CGA

EGA monochrome

During installation of the program approximately 1 Mb of free disk space is required. The installed program (including a data-set with 718 lakes-years x 71 variables) takes 0.7 to 0.9 Mb of disk space, depending on your hardware.

Since some options in the program require heavy calculations, a numerical coprocessor is strongly recommended. Three dimensional plots can only be made automatically when the grid program SURFER is present on your disk (not necessarily in your DOS path). If a mouse is available, it can be used.

There are two versions of BASIS that differ completely in the use of RAM memory. During the installation the program INSTALL automatically selects the right version. The two modes are:

1. *Real mode.* This version runs on any DOS machine with at least 640 Kb memory. It uses expanded memory (EMS) if available.
2. *Protected mode.* This version runs on an 80286 processor or higher, if there is at least 2 MB extended or expanded memory available. It uses all available memory and runs DOS in the 'protected mode'. It is faster than the real mode version, in particular with large data sets.

2.2. Using INSTALL

INSTALL detects what hardware you are using and configurates BASIS appropriately. Files that are not necessary for your machine are deleted.

To install BASIS:

1. Insert the installation disk into drive A. Type the following command and press *[Enter]:*

   A:INSTALL

2. Enter the directory of the source (usually *A:*) and press *[Enter]*.
3. Enter the directory of the destination (usually *C:\BASIS*) and press *[Enter]*.

2.3. Running BASIS

To start BASIS, go to the BASIS directory created with INSTALL. Usually this directory is *C:\BASIS*.
To start BASIS type:

BASIS (followed by *[Enter]*)

Now you enter the input screen, where you can select variables.

2.4. Getting help

Wherever you are in the program, it is always possible to get case-sensitive help. Press *[F1]* to enter the on-line help system. To learn more about BASIS it is possible to get help on related topics and go to an index with ± 85 entries.

1. **Case-sensitive help.**

   Press the *[F1]* key any moment in the program to get case-sensitive help. Use *cursor keys* or the vertical *scroll bar* (see § 2.5) to scroll the help screen.

2. **Help on related topics (mouse click or *[F2]*)**

   In most help screens part of the text is highlighted. These are links. You can use these highlighted links to display a new help screen that presents information about the related topic.

   a. Click on the highlighted words to display that screen.

   b. If you don't use a mouse press *[F2]* to get a list of related topics. Move to the desired bar and press *[Enter]*.

3. **Help index (*[F1]*)**

   If you press *[F1]* in a help screen you get the *Help Index* on your screen. Use the cursor keys or the scroll bar to highlight the right item. Alternative: press the first character of the item. Select the highlighted item by pressing *[Enter]*.

4. **Go to the previous screen (*[Esc]*)**

   If you press *[Esc]*, you go to the previous screen. The help system of BASIS recalls the last 10 topics that are examined, as long as you don't leave the help system.

2.5. Special keys and how to use the mouse

The following keys can be used

[F1]	Help
[F1][F1]	Help index
[F2]	Zoom scatter plot (this key has also other functions, see status bar)
[F3]	Highlight part of a scatter plot.
[F10]	Go from the input screen to the main menu.
[Esc]	Cancel calculation or go to previous menu. All calculation can be cancelled by pressing *[Esc]*. It may, however, take a few seconds before the program reacts!
[Up][Down]	Move cursor or scroll
[PgUp][PgDn]	Scroll up or down one page.
[Home][End]	Go to the first or last page.
[Enter]	Select or change an item
[Alt]+F	Save to or load selection from file

How to use a mouse?

The use of a mouse is fully supported:

The highlighted keys (e.g., *[F1]*) of the status line below can be activated by a mouse click on the key name or on the text after the key name.

In menu's a single mouse click selects an item. In the input screen a single mouse click moves only the cursor. A double click is required to select a value or to enter a new value or choice.

In some menu's vertical scroll bars are displayed (Fig. 1.). You can use this scroll bars with a mouse to scroll the contents of the menu:

1. To scroll one line at a time, click the arrow at either end of the bar.
2. To scroll continuously, keep the mouse button pressed.
3. To scroll on page at a time, click the bar to either side of the scroll box.
4. To move quickly, drag the scroll box to any spot on the scroll bar.

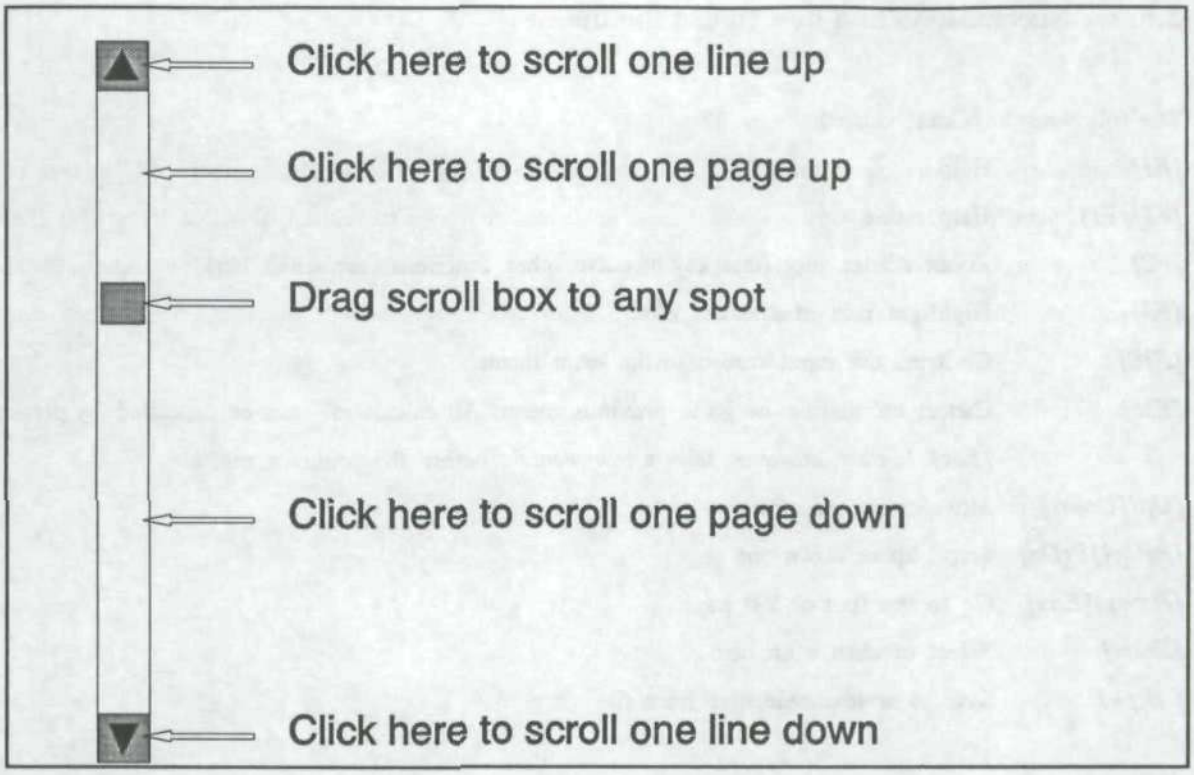

Click here to scroll one line up
Click here to scroll one page up
Drag scroll box to any spot
Click here to scroll one page down
Click here to scroll one line down

Fig. 1. How to use vertical scroll bars

Special use of the mouse in scatter plots

In some scatter plots it is possible to get more information about the points that are displayed. Just a mouse click on the point gives the name of the lake that is displayed. More information about that lake can be obtained by pressing *[F2]* afterwards (see also § 4.5).

3. METHODS

The main purpose of the program BASIS is to find analogous cases based on available information. In chapter 3.1 is explained how this is done. Based on these analogous cases it is possible to predict the response of the question lake. In chapter 3.2 the averaging method is explained. The next question is how good this prediction is. In chapter 3.3 is explained how the percentage of variance explained can be calculated. The methods and parameters used by BASIS can be optimized automatically. Chapter 3.4 explains by which method this is done.

3.1. How to find analogies?

The steps to be taken to find analogies are summarized in Fig. 2a. Each step is explained below.

Select variables of question lake

The first step in Weighted Analogies Prediction is to select variables that are to be used in the analysis. There are two kinds of variables:

1. *Conditional variables.*
   Conditional variables are called independent variables in regression analysis. They 'explain' the state of the ecosystem of a lake. These variables are known or can be managed in the question lake.
2. *Response variables.*
   Response variables are variables that express the state of the ecosystem of a lake. In regression these variables are called dependent variables.

Class variables

Class variables need special attention, since these variables are not fully quantitative. We distinguish two kinds of class variables: nominal and ordinal variables:

1. Values assigned to the classes of *nominal variables* do not imply any order. Differences between them, have no function. Examples: sediment type, lake manager.
2. Values assigned to the classes of *ordinal variables* imply an order. Differences, however, have no absolute meaning. Example: Retention time classes.

Fig 2a. How to predict?

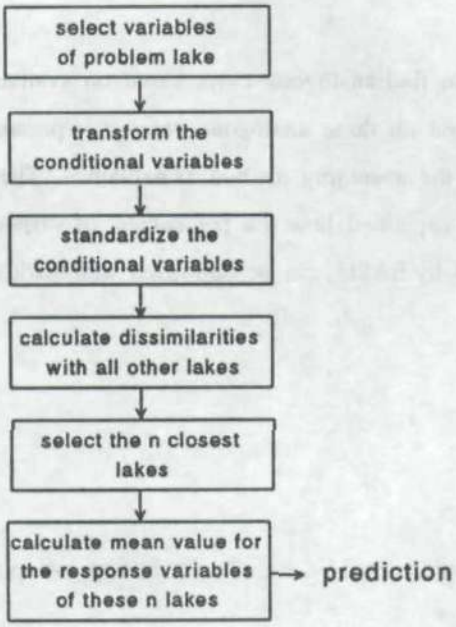

select variables of problem lake
transform the conditional variables
standardize the conditional variables
calculate dissimilarities with all other lakes
select the n closest lakes
calculate mean value for the response variables of these n lakes
prediction

Fig. 2b. Weighted Analogies Prediction

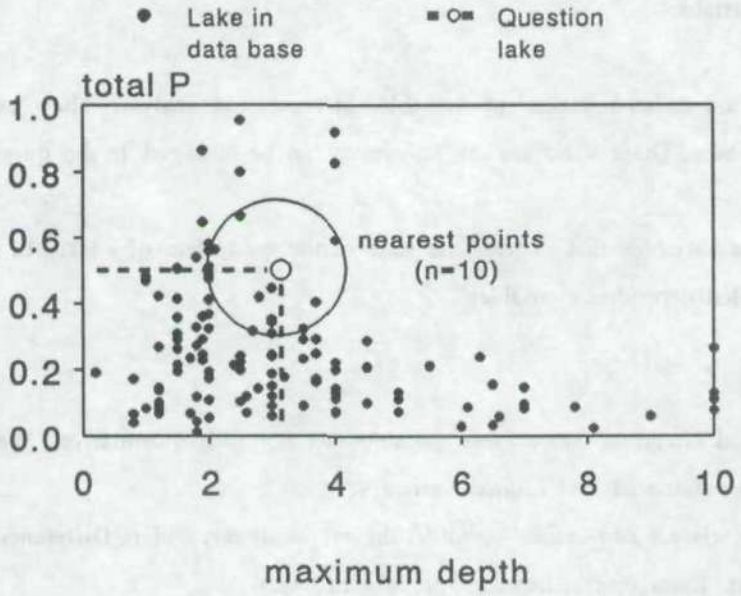


Fig. 2. (a) Schedule of the method of predictions (b) Example

In BASIS nominal variables may not be used as response variables, because the mean of a nominal variable lacks meaning.

When nominal variables are used as conditional variables, the variable is translated into binary dummy variables. Each class is one variable that can take two values: true (1) or false (0).

For ordinary variables these restrictions are not made. The results of predictions of these variables still may be not very accurate.

Transformations

Before standardization, a non-linear transformation can be used to shrunk certain parts of the scale and to stretch other parts. The most common used transformation is the logarithmic transformation.

The following transformations and standardization are available:

1. *Logarithmic transformation*

   Commonly used transformation to change a log-normally distributed variable into a normal distribution or to give less weight to large quantities.

$$y_{ki}^{*}=\ln(y_{ki}+1)$$

2. *Log(x/(100-x)) transformation*

   Special transformation to bring data of percentages of phytoplankton as close as possible to a normal distribution. See: Trimbee & Prepas (1987)

$$y_{ki}^{*}=\ln(\frac{y_{ki}}{100-y_{ki}})$$

3. *Square root transformation*

   Square root transformation, used before analysis of Poisson-distributed variables.

$$y_{ki}^{*}=\sqrt{y_{ki}}$$

4. *Inverse transformation*

Used, e.g., to make the relation between Secchi depth and extinction approximately linear:

$$y_{ki}^{*}=\frac{1}{y_{ki}}$$

5. *No transformation*

Use untransformed data.

Symbols:

y_{ki} value of the variable k in lake i

n number of variables

Standardization

The different variables need to be standardized to give equal weight to different scaled variables.

1. *Normalization*

By this method the relative position of an observation within a distribution is described. The normalized value (also called *standard* or Z score (SPSS, 1987)), shows how many times the standard deviation an observation deviates from the mean of the population. The mean of the normalized values is 0 and the standard deviation is 1. It is obtained by subtracting the mean from a value and dividing this difference by the standard deviation.

$$y_{ki}^{*}=\frac{y_{ki}-\overline{y_k}}{s.d.}$$

2. *MinMax standardization*

By this method the variables are scaled between the minimum and maximum values in the database. The minimum gets value 0 and the maximum gets value 1.

$$y_{ki}^{*}=\frac{y_{ki}-MIN_i(y_{ki})}{MAX_i(y_{ki})-MIN_i(y_{ki})}$$

5. *No* Use unstandardized data. Use this option only if the data are already standardized.

Symbols:

y^*_{ki}	transformed value of the variable k in lake i
y_{ki}	value of the variable k in lake i
$MIN_i(y_{ki})$	minimal value of the variable k in the database
$MAX_i(y_{ki})$	maximal value of the variable k in the database
$\hat{y}_k$	mean of variable k in the database
s.d.	standard deviation of variable k in the database
n	number of variables

Weighting the variables

The variables are weighted by multiplying their values with weights that are entered by the user. The purpose of this weighting is to give important variables more weight. The weights also can be optimized by the computer (see § 3.4).

Calculate dissimilarities

The distance between the 'question lake' and all other lakes is calculated by a dissimilarity index.

Supported indices:

1. *Euclidean distance* (default). The Euclidean Distance (ED) is the most frequently used index. It is the distance in the n-dimensional space constrained by the conditional variables (Each variable is one dimension of the space).

$$ED=\sqrt{\sum_{k=1}^{n}(y_{ki}-y_{kj})^2}$$

2. *City Block distance* (also called the Manhattan Distance). The City Block Distance (CB) is the sum of the differences between all variables. It weights the variables that are far out stronger than the Euclidean Distance does.

$$CB=\sum_{k=1}^{n}|y_{ki}-y_{kj}|$$

3. *Cord Distance*. The Cord Distance (CD) is geometrically represented by the distance between points where the sample vectors intersect a unit sphere (see: Jongman *et al.*, 1987). It gives more weight to qualitative aspects than the other indices of the program.

$$CD=\sqrt{\sum_{k=1}^{n}\left[\frac{y_{ki}}{\sqrt{\sum_{k=1}^{n}y_{ki}^2}}-\frac{y_{kj}}{\sqrt{\sum_{k=1}^{n}y_{kj}^2}}\right]^2}$$

4. *Chebychev Distance*. The Chebychev Distance (ChD) is the maximum difference between variables. It weights one variable that is far out even stronger than the City Block Distance.

$$ChD=MAX_k|y_{ki}-y_{kj}|$$

Symbols:

y_{ki} transformed and standardized value of the variable k in lake i multiplied with the weight of the variable (default weights are 1).

n number of variables

The properties of the Euclidean Distance and the Cord Distance are discussed by Jongman, *et al.*, (1987).

3.2. How to predict a response variable?

After calculation of the dissimilarities of all lakes with the question lake, the lakes in the database are ranked according to the obtained values. The N nearest lakes are used to make a prediction of the selected response variables (see Fig. 2b) (we call N the *number of nearest points*). The default number for N is 25. The response variables of these cases are weighted so that the influence of the cases declines with the dissimilarity from the lake being estimated.

$$P^*=\frac{\sum_{i=1}^{N}(y^*_{ki}\ D_i^p)}{\sum_{i=1}^{N}D_i^p}$$

In which:

P^* Prediction of the transformed response variable (needs to be transformed back by using the inverse of the transformation)

N Number of nearest points

y_{ki} transformed (not standardized) response variable k of lake i

D_i Dissimilarity of lake i with the question lake

p Distance weighting power (is always negative !)

The more negative the *distance weighting power*, the faster the decline in influence and the less the effect of points further out will have on the interpolation.

3.3. How accurate is the prediction?

To answer the question how accurate the prediction is, a special technique was used, called *cross validation* (also called 'All-but-one-method'). This is a data efficient way of testing a model (see also Fig. 3a.). Each time, one lake from the database is removed and treated as the unknown case. The rest of the lakes is used to make a prediction using Weighted Analogies Prediction. This procedure is repeated until all lakes in the database are predicted. The predictions are compared with the observed values. The values of observed and predicted are plotted (see Fig. 3b.), and the percentage of the variance (R^2) that is explained by the model is calculated.

R^2 (sometimes called coefficient of determination) is the same statistic that is commonly used in linear regression. The sample R^2 usually is an optimistic estimate of how well the model fits the reality. The statistic adjusted R^2 attempts to correct R^2 to reflect more closely the goodness of fit of the model in the population.

Formula:

$$adj.R^2 = 1 - \frac{residual\ sum\ of\ squares/(N-p-1)}{total\ sum\ of\ squares/(N-1)}$$

N = sample size, p = number of parameters.

If the adjusted R^2 equals 1, the model fits perfectly, if the adjusted R^2 is negative, the mean value of the response variables is a better prediction than the model.

Fig. 3a. Cross validation

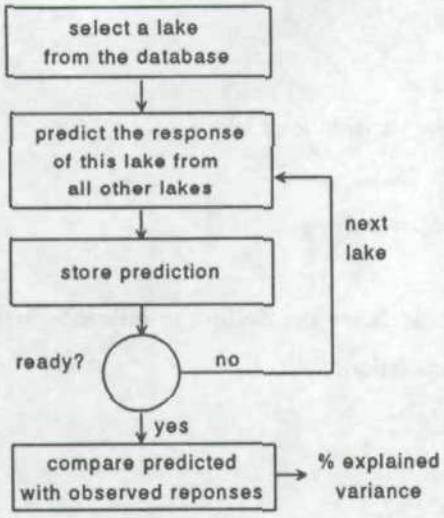


Fig. 3b. Goodness of fit of WAP
adj.R^2=0.46 (regression: 0.29)

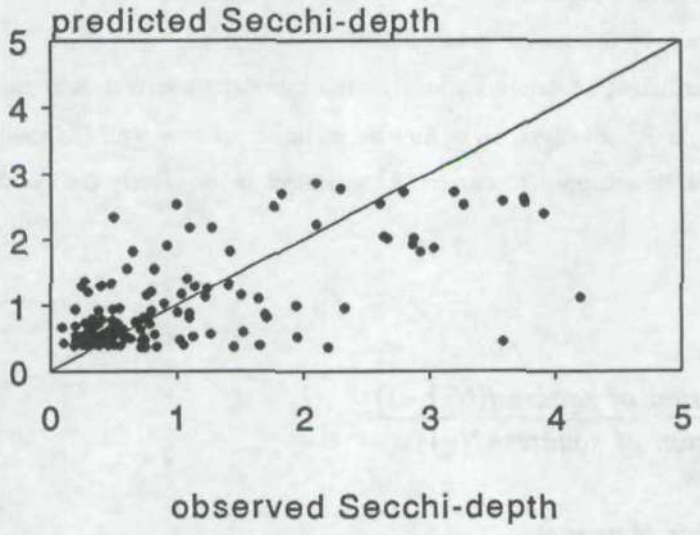


Fig. 3. (a) Schedule of the method of cross validation (b) Example

3.4. Optimization of the prediction

To help the user to set the values of parameters like the weights of the conditional variables, an optimization routine is included in the program.

The following parameters used by the prediction method can be optimized automatically:

- *Weights of conditional variables*
- *Distance weighting power*
- *Number of nearest points*

The parameters are optimized iteratively, by use of the *Controlled Random Search (CRS)* algorithm (de Hoop *et al.*, 1992, Klepper *et al.*, subm.). This algorithm is an improvement of pure random search, an algorithm searching the best set of parameters by trying at random. After each iteration the goodness of fit (adjusted R^2) is calculated by cross validation (see § 3.3).

The CRS algorithm first selects N sets of model parameters uniformly distributed over prior parameter ranges, calculates the goodness-of-fit for each and puts them in a vase.
It then selects m+1 points at random from the vase and mirrors the last point over the average (centroid) of the first m. The mirrored point is the new trial point. The goodness-of-fit (R^2) is calculated. If the R^2 better than the worst set of parameters in the vase, the worst element of the vase is replaced by this new guess. This process continues until a convergence criterion is reached.

An option is to add all variables stepwise. Each step is followed by an iterative optimization.

4. TUTORIAL: AN EXAMPLE OF A PREDICTION

In the following exercises, we will make a simple prediction and optimize the results. We try to predict the summer chlorophyll-a concentrations and the Secchi-depth of Lough Neagh (Sas, 1989) using the mean depth (8.9 m) and the total-P concentration in the summer 1989 (0.106 mg/l). The summer mean Secchi and chlorophyll values were respectively 0.85 m and 79 µg/l (Sas, 1989).

4.1. Selecting your variables

In the first exercise we'll start the program from DOS and select the conditional (independent) and the response (dependent) variables. How to get on-line help from the program is explained too.

To start the program and get some help:

1. Enter in DOS:

   BASIS (followed by *[Enter]*)

   The program displays a message about restrictions in the use of the data-set.
2. Press any key to enter the input screen (see Fig. 4). The first column shows codes for variables. The full names of the variables are shown below in the status bar (only if the cursor is on the first column).

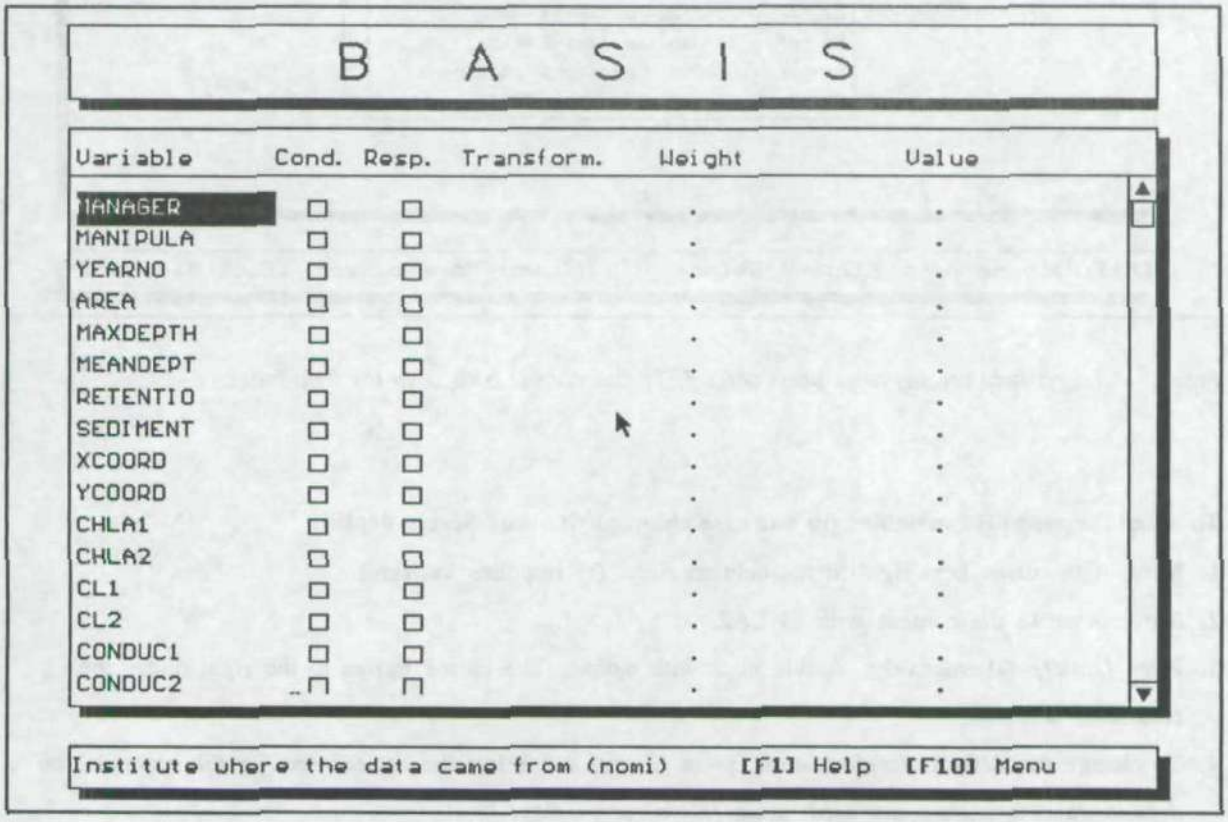


Fig. 4. The input screen.

3. To get help about the column *Cond.* press the right arrow key once and press *[F1]*.
   Now a help screen about *Conditional variables* is displayed. The highlighted words are related topics.
4. To get information about a related topic (e.g., *the input screen*) click on the highlighted word (*input screen*). Alternatively: press *[F2]* to display a menu with related topics and select *input screen*.
5. Press *[Esc]* to go back to the first help screen.
6. Press the *[F1]*-key. Now a full list of topics is displayed (see: Fig. 5).
7. Scroll the list of topics (cursor keys or scroll bar) and press *[Enter]* (or click) on an item. (Alternatively press the first character to move faster to the right spot)
8. Press *space bar* to exit the help system.

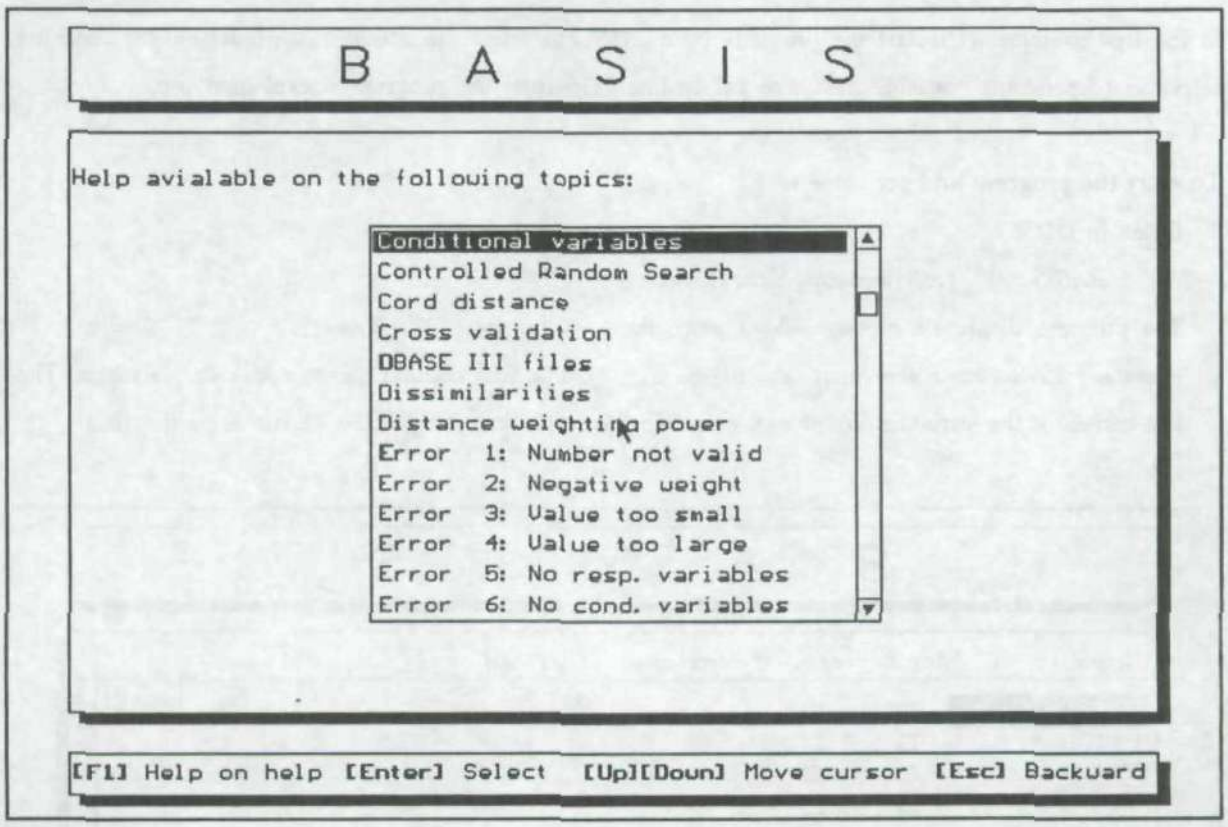

B A S I S
Help avialable on the following topics:
Conditional variables
Controlled Random Search
Cord distance
Cross validation
DBASE III files
Dissimilarities
Distance weighting power
Error 1: Number not valid
Error 2: Negative weight
Error 3: Value too small
Error 4: Value too large
Error 5: No resp. variables
Error 6: No cond. variables
[F1] Help on help [Enter] Select [Up][Down] Move cursor [Esc] Backward

Fig. 5. To get help on any topic press twice [F1] and choose a topic in the help index.

To enter the response variables (in our case chlorophyll a and Secchi-depth):

1. Move with cursor keys right to the column *Resp.* (= response variable).
2. Move down to the column with CHLA2.
3. Press *[Enter]*. (alternatively: double click with mouse) The cursor moves to the right under the column *Transform*.
4. To change the default transformation press *[Enter]* and select the desired one. In this example the default transformations are used: press *[Right]* arrow key.
5. Do the same with SECCHI2.

To enter the conditional variables:

1. Move to the column *Cond.* (= conditional variable) and position the cursor on the right row (MEANDEPT).
2. Press *[Enter]* or double click with mouse. The cursor moves to the column *Tran sform.* We don't change the default transformation, so press to the *[Right]* arrow key.
3. Press the *[Right]* arrow key.

   To change the default weight (1.0) to be given to the variable press *[Enter]*, else move to the right.
4. Press *[Enter]* to enter the value of the question lake (in our case: *8.9*) followed by *[Enter]*.
5. Now enter the second conditional variable in the same way (*PTOT2*, transformation: *Logarithmic*, weight: *1.0* and value = *0.106*)
6. After selection of the conditional variable the screen looks like Fig. 6.

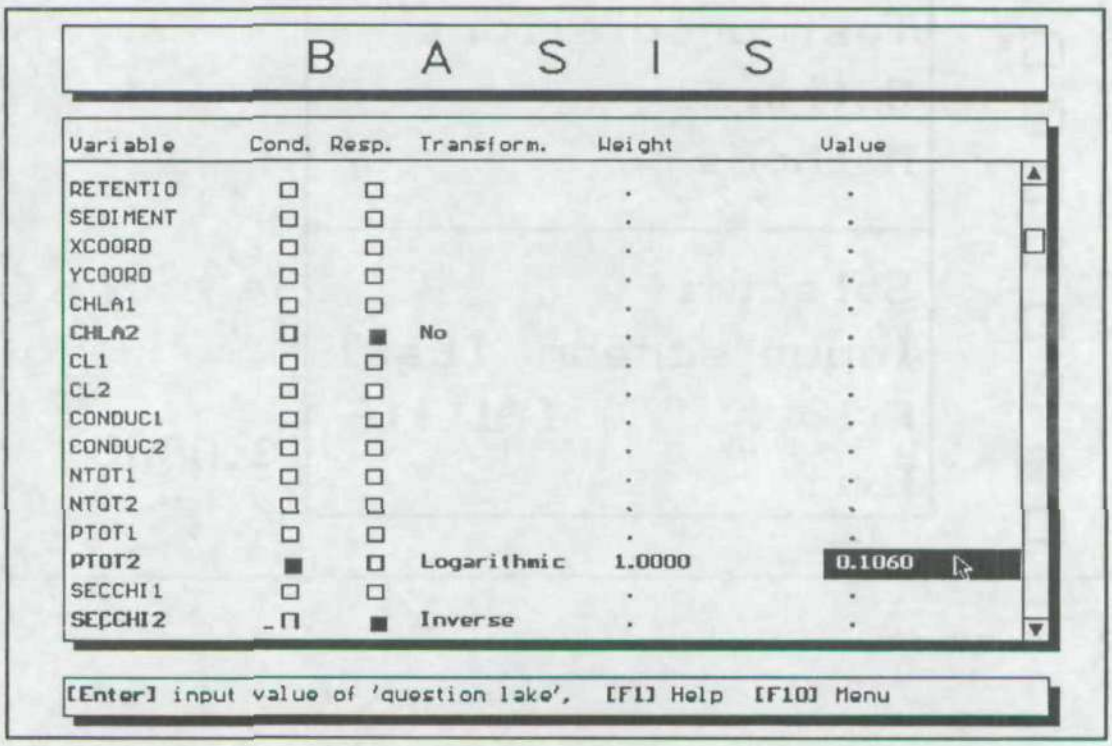


Fig. 6. After selection of the variables the screen looks like this.

4.2. The main menu

After selection of the desired variables, different kinds of analysis can be selected in the main menu.

To open the menu from the input screen:

1. Press *[F10]*.

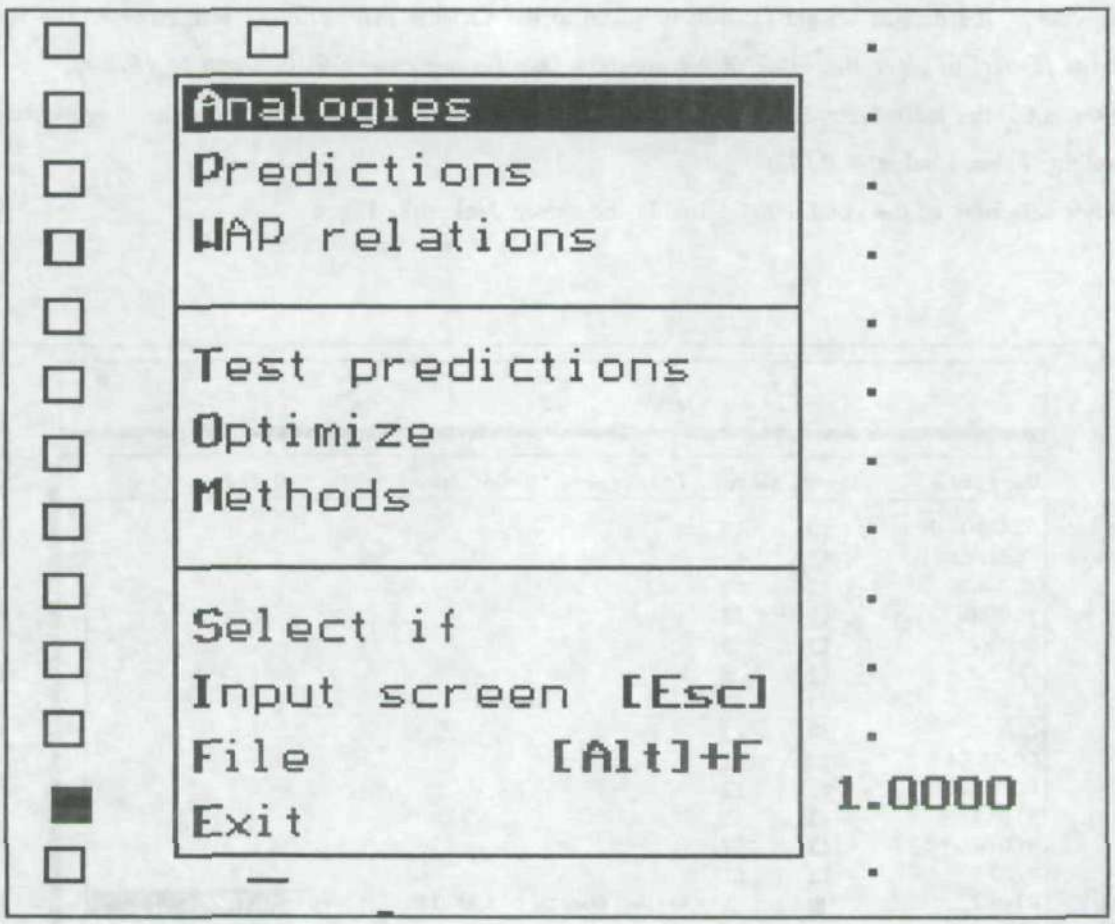


Fig. 7. The main menu.

A short description of the items in the main menu:

Run Weighted Average Prediction:

Analogies

Find lakes that resemble the question lake

Predictions

Predict the response variables of the question lake

WAP relations

Show relations between response variables and conditional variables.

Optimize methods:

Test predictions

Test the prediction model and calculate the adjusted R^2 for each response variable. This is done by cross validation.

Optimize

Optimize weights and/or methods automatically by an iterative algorithm. Warning: this takes some time..

Methods

Change methods of prediction and dissimilarity indices.

Selection of data or variables:

Select if

Enter or change a condition for the database (for example: *sediment='Sand'*)

Input screen

Return to the selection of parameters

File

Save or load selection to a file.

Exit

Return to DOS.

4.3. Filtering the database

In this exercise we will select from the data base only those lakes that are not recently managed and that have a maximum depth of less than 10 meter. All other lakes are not used in the analyses. To make such selection a logical expression is used. A logical expression has one or more relational expressions that are combined by logical operators.

A relational expression consists of:

1. a variable name
2. a relational operator:
   - = equal
   - <> not equal
   - > greater than
   - < less than
   - >= greater than or equal to
   - <= less than or equal to
3. a value or class name. Class names (e.g., 'Sand') must be written between delimiters (').

An example of a simple relational expression.

(SEDIMENT < > 'Sand')	To select all data from the database except the lakes with sandy sediment.
PTOT1 > = 0.1	To select all lakes with tot-P of more than 0.1.

Relational expressions can be combined by logical operators. Parentheses () must be used in combination with the logical operators!

The logical operators are (see also Table 1):

not	negation
and	logical and: to combine two conditions
or	logical or: to combine two conditions
xor	logical exclusive or: to combine two conditions

Examples.

(Sediment = 'Sand') and ((PTOT1 < 0.2) or (PTOT1 > 0.1))

not (sediment = 'Sand') (the same as: Sediment < > 'Sand')

Table 1. Logical operators are used to combine to relational equations that can be TRUE or FALSE. How these operators work is explained in this table.

suppose A is	TRUE	FALSE	TRUE	FALSE
and B is	TRUE	TRUE	FALSE	FALSE
than A and B	TRUE	FALSE	FALSE	FALSE
A or B	TRUE	TRUE	TRUE	FALSE
A xor B	FALSE	TRUE	TRUE	FALSE
not A	FALSE	TRUE	FALSE	TRUE

To select all lakes with a maximal depth of less or equal than 10 m, except manipulated lakes.

1. Select the bar *Select if* in the main menu and press *[Enter]* (or click *Select if*).
   Now a dialog screen is displayed. You can either use the menu or the keyboard to enter an expression. We will use the menu.
2. Select *(left parenthesis* (and press *[Enter])*.
3. Select *Select variable*.
   Now a list of variables is displayed.
4. Select *MANIPULA* from the list of variables and press *[Enter]* or double click.
5. Select the greater than sign (>).

Oops: wrong symbol!

6. Press *[Backspace]* to clear last symbol (>).
7. Select the equal sign (=).
8. Select *Select class*.
9. Select *'No'* from the list of class names and press *[Enter]*.
9. Select the *) right parenthesis*.
10. Select *AND*.
2. Select *(left parenthesis*.
11. Select *Select variable*.

    Now a list of variables is displayed.
9. Select *MAXDEPTH* from the list of variables and press *[Enter]* or double click.
12. Select the less equal sign (<=).
13. Enter *10*.
14. Select the *) right parenthesis*.

    Now the condition is completed (see: Fig. 8)
15. Select *Ready* to save the expression

    Note: if you press *[Esc]*, the default expression remains unchanged.

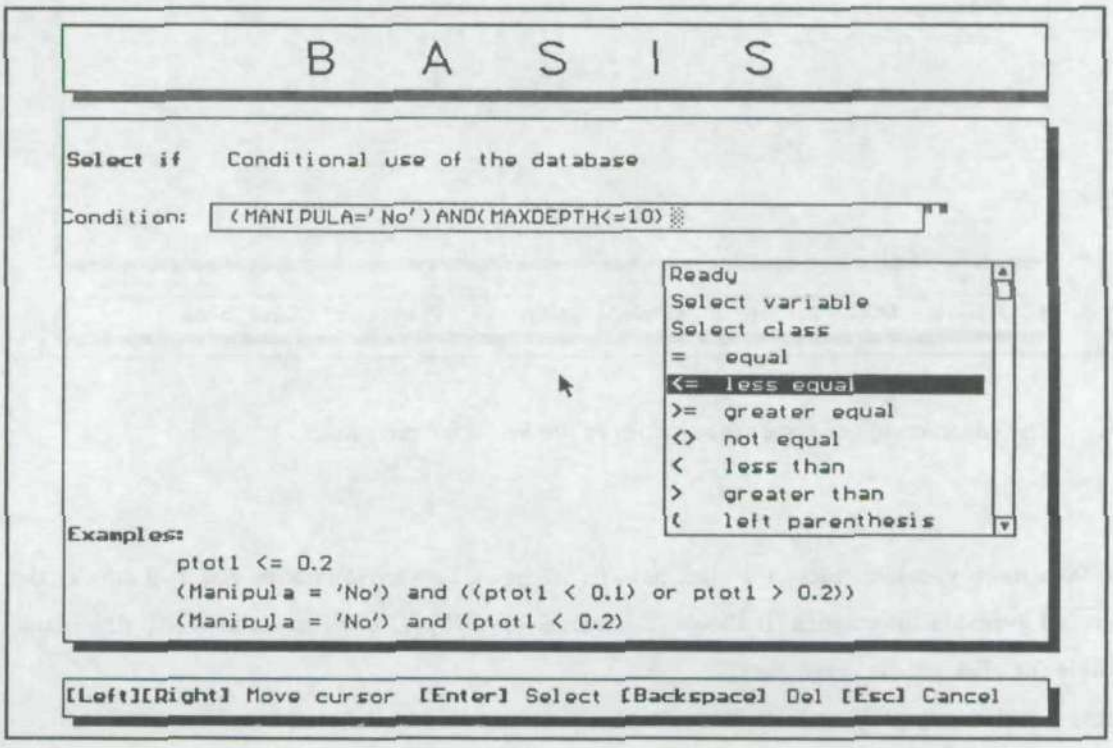


Fig. 8. A logical expression can be entered to select part of the database.

4.4. Finding analogies

In this exercise we shown how analogous lakes can be found. It is possible to browse through the database starting with the best analogous lake.

To find analogies:

1. Select the bar *Analogies* in the main menu and press *[Enter]* (or click *Analogies*).
   The variables of best analogous case are shown on your screen (Fig. 9). Only the selected conditional and response variables are shown.

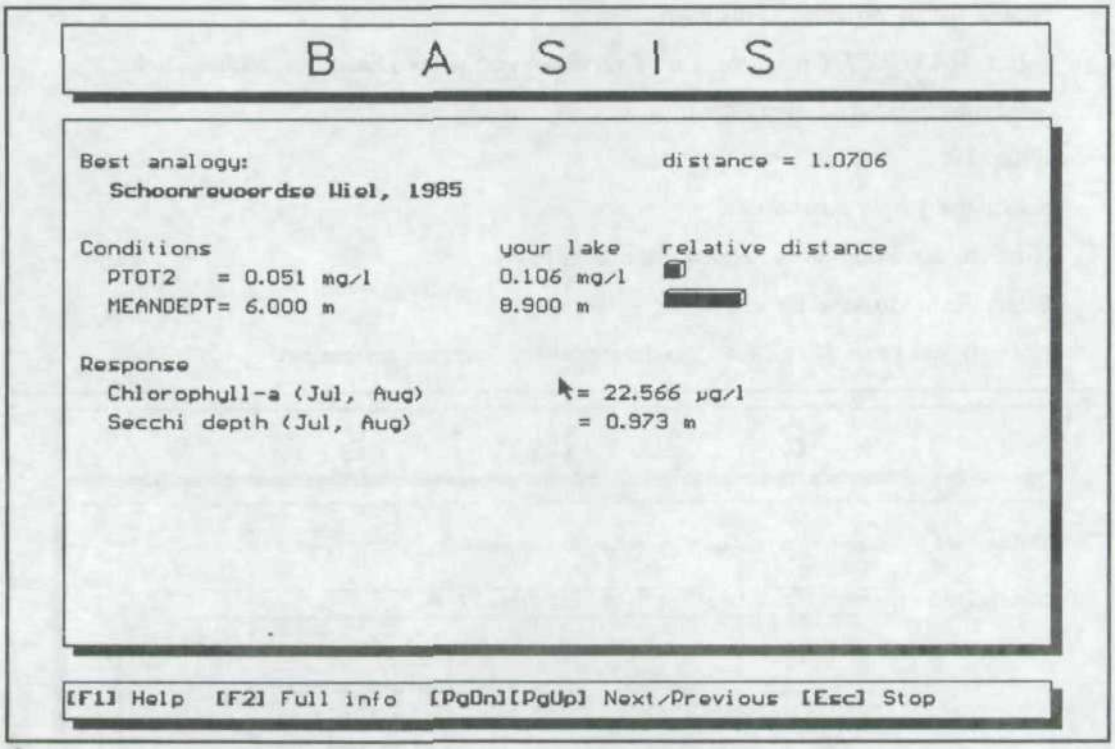


Fig. 9. The conditional and response variables of the best analogous lake.

2. To view more variables: press (or click on) *[F2]* (toggles between *Summary* and *Full info.* mode). Now all available information is shown. Use cursor keys *[Up][Down]* to scroll if not all variables are visible (or click on the scroll bar).
3. Press (or click on) *[F2]* again to return to the *summary* mode.
4. Press (or click on) *[PgDn]* to view other analogies in descending order of similarity.
5. Press (or click on) *[Esc]* to return to the main menu

4.5. Predicting a response

In this exercise we predict the response of Loch Neach. Further we learn how to zoom in scatter plots, how to highlight a part of a plot and how individual points of the plot can be examined. These options also can be used for two dimensional WAP relations (§ 4.6.).

To make a prediction:

1. Select the bar *Predictions* in the main menu to make a prediction. A scatter plot is drawn on your screen (Fig. 10.):

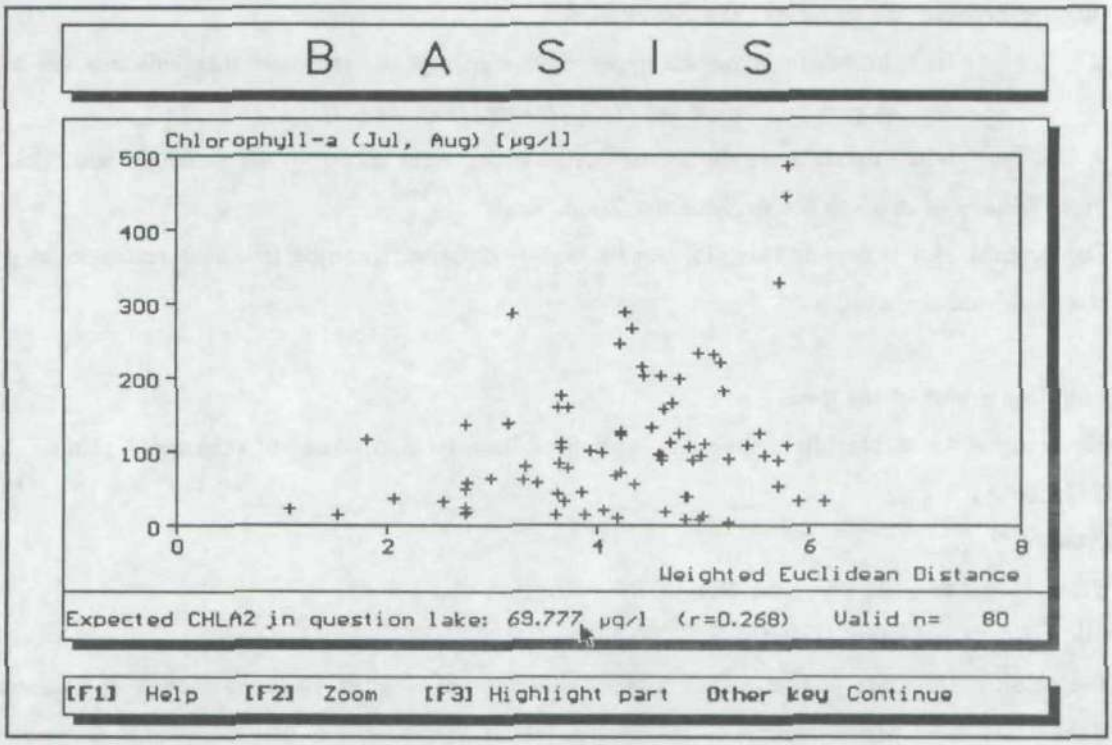


Fig. 10. Plot of the prediction. The predicted chlorophyll is 39 µg/l.

The prediction is visualized by plotting of each lake in the database the response variable against the dissimilarity with the question lake. When the prediction makes sense, a converging cloud of points is expected, with the lowest dispersion near the y axis. Under the figure the prediction the question lake is displayed.

The *r* that is displayed is a measure of how well the cloud converges. It is the correlation between dissimilarity and difference between response variable and prediction. So: near 1 = perfect converging cloud near -1 = diverging.

2. Press *[Enter]* to view the prediction of the next response variable (SECCHI2).

Zooming in:

1. Press *[F2]* to enter the *zoom mode*.
   The selected area is indicated by a rectangle with the blinking red cross in the right upper corner. You will find this rectangle in the lower left corner of the scatter plot.
2. To change the *Size* of the rectangle:
   a. *Without mouse:* move the upper right corner of the rectangle that indicates the selected area by pressing cursor keys (*[Up][Down][Left][Right]*).
   b. *With mouse:* move the mouse to the upper right corner of the rectangle and click.
3. To *Move* the rectangle press *[F2]*
   (toggles between the *move* and the *Resize* Mode).
   a. *Without mouse:* move the upper right corner of the rectangle that indicates the selected area by pressing cursor keys (*[Up][Down][Left][Right]*).
   b. *With mouse:* move the mouse to the upper right corner of the rectangle and click.
4. Press *[Enter]* or *double click* to leave the Zoom mode.
   The zoomed plot is drawn. This plot can be slightly different from the indicated rectangle to get nice scales on the axes.

Highlighting a part of the data:

In this example we will highlight the lakes with more than 10 % coverage of submerged plants (SUBMERG > 10)

1. Press *[F3]*
2. Enter a condition (in our case: *SUBMERG>10*). (see also § 4.3.)
3. Select *Ready* and press *[Enter]*
   The selected lakes are plotted as red circles, while the not selected lakes are plotted as blue crosses. If there are cases where there is no information about the vegetation they are plotted as gray crosses. Unless this selection is changed, this selection is used in all plots (predictions and WAP relations) during the BASIS run.
4. Press *[F1]* to display a legend, press *[Esc]* to return to the plot.

Examine individual points (mouse only):

1. Click on a point to see the name of the lake and the year of the data (see: Fig. 11).
2. Click on the lake name (or press *[F2]*) to get a full list of information about the lake.
3. Scroll the text if necessary.
3. Press *Other key* to return to the plot.

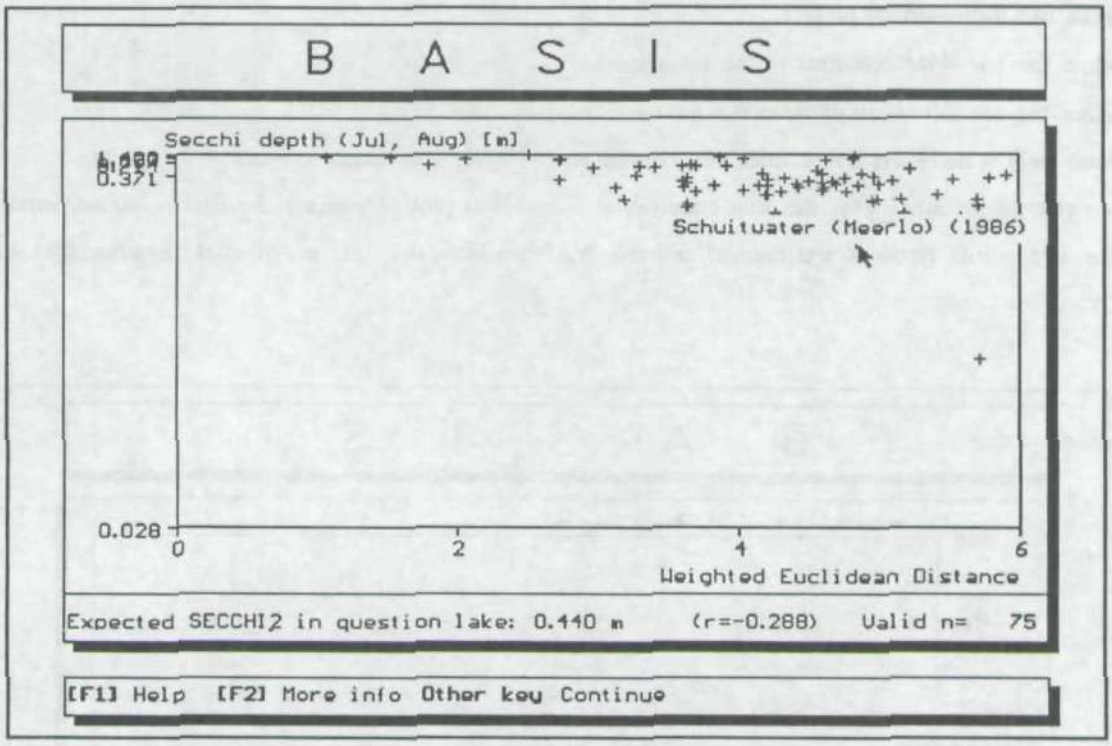


Fig. 11. Click to examine individual points

4. Click between points.

   Now the mouse cursor jumps to the nearest point.

5. Click again to remove the text from the plot.
6. Press *[Space]* until you are returned to the main menu.

4.6. Viewing relations between variables

In this exercise we learn about the option for exploratory analysis of the data set. Two dimensional scatter plots and three dimensional surface plots show the relationships between the conditional and the response variables.

The mean relationships between the variables are shown by making WAP predictions. The technique is than comparable to the 'running averages' method. The way of averaging of WAP (§ 3.2), however, differs and can be optimized.

Note: Only the visible conditional variables are used in the calculations.

To make two dimensional plots:

1. Select the bar *WAP relations* in the main menu
2. Select the bar *2D relations* from the next menu.

   Note: only if there are more than one conditional variable this menu appears.

   In a two dimensional plot the first conditional variable is plotted against the first response variable. The data points (crosses and circles) and the WAP predictions (line) are plotted together (see Fig. 12).

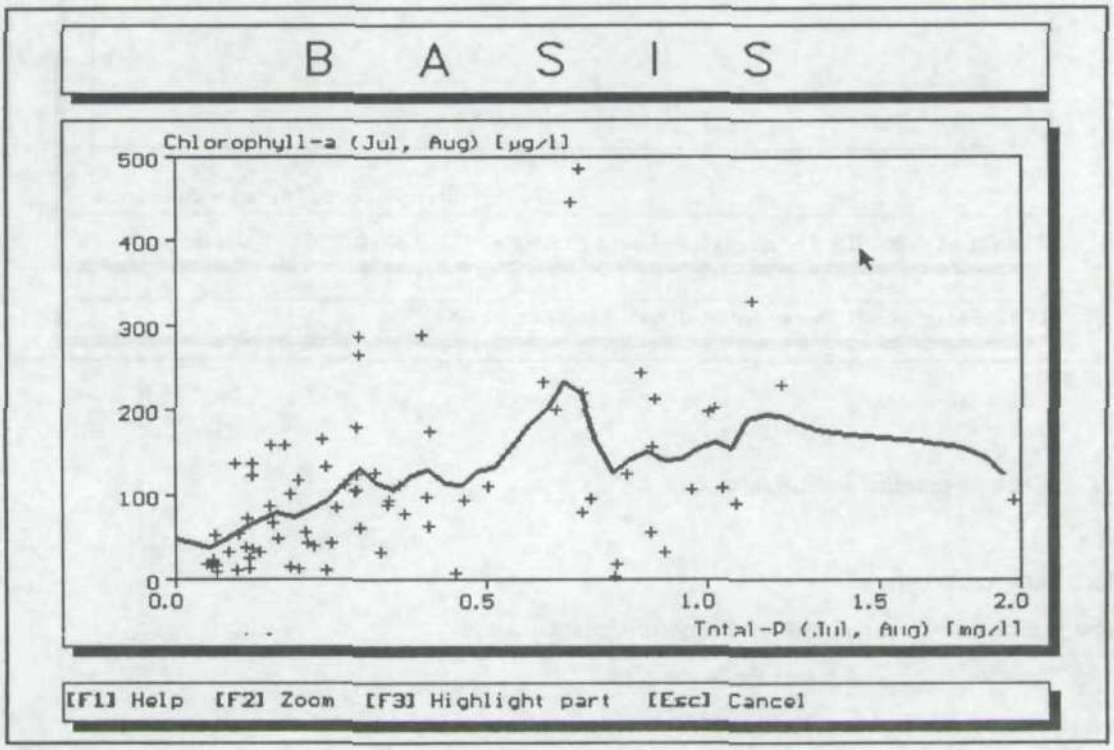


Fig. 12. Two dimensional plot with the WAP predictions.

In this plot it is also possible to:

- Zoom in
- Highlight part of the plot
- Examine individual points (mouse only)

In the previous exercise we have learned how to do these things.

4. Press *[Esc]* to view next plot, until all combinations of conditional and response variables are shown. Then you return to the mean menu.

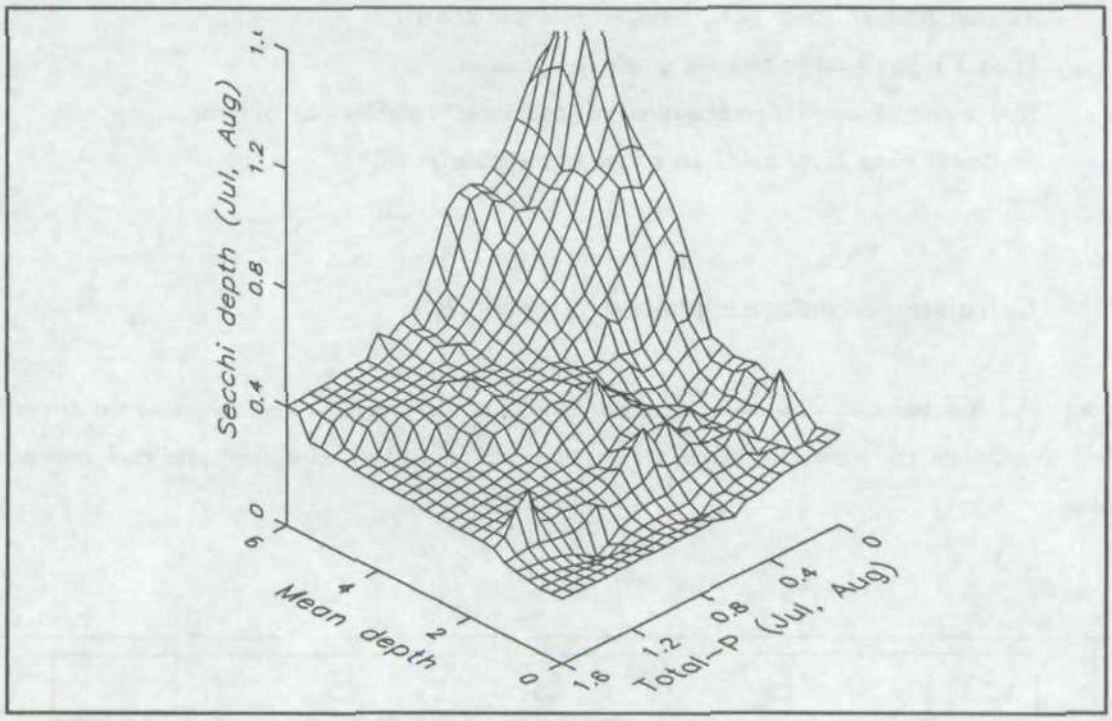


Fig. 13. Three dimensional plot of two conditional variables and one response variable.

To make three dimensional surface plots:

1. Select the bar *WAP relations* in the main menu
2. Select the bar *3D relations* from the next menu.
3. If you don't have SURFER on your current drive, the following error message is displayed:
   Error 14: SURFER not found
   If this message appears, it is not possible to make 3D plots at run time. The ASCII grid files still can be made, press *[Enter]* and follow the next instructions to do so. If you want to cancel the calculations: press *[Enter]* and *[Esc]*.
4. Now a two dimensional plot of the conditional variables is shown. Press *[F2]* if you want to select a certain part of the plot. You should do this to exclude parts of the plot with few data, to avoid large interpolations. You enter the *zoom mode* of the plot. You have learned in § 4.5 how to zoom in.
5. Press *[Enter]* to start the calculations.
   All grid points that are calculated are displayed, until the whole plot is filled. Then SURFER is started and the three dimensional plot is drawn (Fig. 13.).
6. Press *[Esc]* to go on.
   The next question appears:
   Enter rotation of previous plot >
7. Enter *30* followed by *[Enter]*.
8. Now you return to the 3D plot, which is rotated 30 degrees.
9. Press *[Esc]* twice to rotate another 30 degrees.

10. To stop rotation, press *[Esc]* followed by *0* and *[Enter]*
11. Press *Y* if you want to plot the graph else press *N*.
12. Now a plot of the next combination of conditional variables can be made.
    To cancel press *[Esc]* until you return to the main menu.

4.7. Calculate percentage explained variance (R^2)

The next two exercises show the more advanced options of the program. First the cross validation is applied to calculate the percentage explained variance and plot observed against predicted response variables.

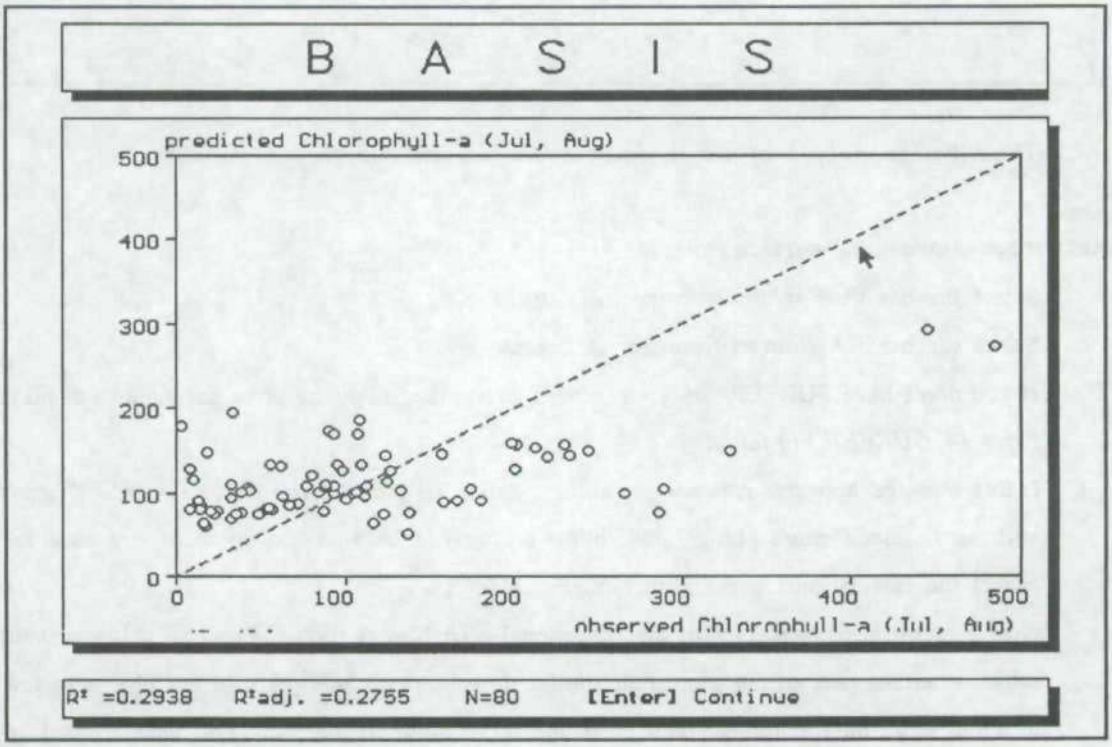


Fig. 14. Test predictions and calculate the percentage explained variance (adjusted R^2).

To test the predictions (takes some calculation time):

1. Select the bar *Test prediction* in the main menu
   Slowly Fig. 14 is build on your screen. Press *[Esc]* to cancel, otherwise wait until the adjusted R^2 is displayed.
2. Press *any key* to view the next relation.
3. Repeat this until all response variables are tested and the main menu reappears.

4.8. Optimizing your model

In this exercise we optimize the methods and the weights of the conditional variables. The calculations take much time, so use a fast computer, and let it work for a night or weekend.

To optimize your model:

1. Select *Optimize* in the main menu.
2. Select *Enter all variables* in the next menu.

   If you have many conditional variables and you want to select the best explaining variables, select *Stepwise*. Then the variables are selected stepwise, with optionally an iteration after each step.
3. Select *Optimize full* and wait. After some time a plot of the weights of the prediction exponent and the adjusted R^2 is drawn on the screen. After each iteration a point is drawn in the plot. (Fig. 15).
4. If you want to cancel the calculations, press *[Esc]*. The best model sofar is then stored.
5. After a convergence criterion is reached, an iteration report is written on the screen.

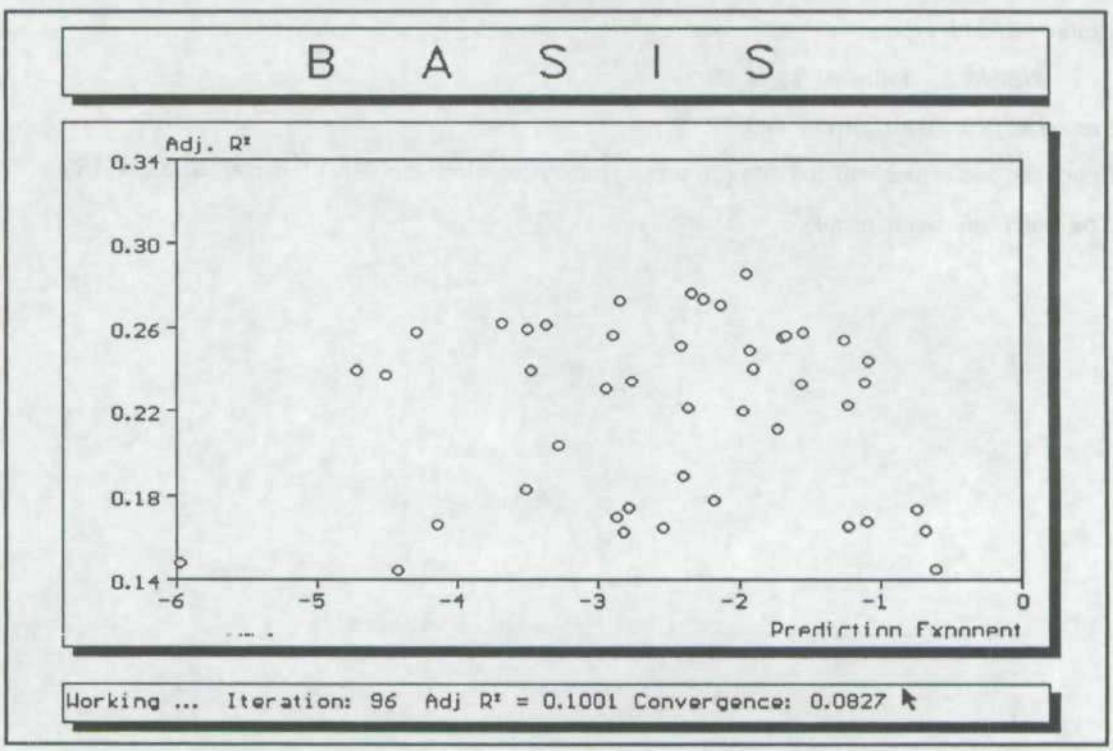


Fig. 15. Optimization of weights.

4.9. Saving and loading your model

In this last exercise is shown how all selections can be saved. Not only the selection of the variables will be saved, but also the logical expressions (to select part of the database or to highlight points in plots) that are entered and the used methods are saved.

To save your optimized model:

1. Select the bar *File* in the main menu.
2. Select the bar *Save to file* in the next menu.
3. Enter the name of the file (or press *[Enter]* to select the default name: *BASIS.VAR*).
   You go back to the main menu.

To exit BASIS and restart BASIS with the same selection:

1. Select the bar *Exit* in the main menu.
2. Restart BASIS typing:
   BASIS (followed by *[Enter]*)
3. Press *[Alt]+F* (both alt-key and F).
4. Enter the same name of the file (or press *[Enter]* to select the default name: *BASIS.VAR*).
   You enter the main menu.

5. GLOSSARY

Adjusted R^2

The statistic adjusted R^2 is the proportion of variance explained by a model. It attempts to correct R^2 to reflect more closely the goodness of fit of the model in the population.

Analogies

Analogies are lakes that resemble the question lake most closely.

BASIS

Best Analogous Information System, a computer system to search for information about analogous cases.

Classes

Nominal or ordinal variables have classes. These variables can only have some discrete values. Examples: Sediment type, country, manager.

Conditional variables

Conditional variables are called independent variables in regression analysis. They 'explain' the state of the ecosystem of a lake.

Examples of conditional variables in limnology:

Mean depth	Maximum depth
Area	Nutrients (tot-P, tot-N)
Chloride concentration	Sediment class

Controlled Random Search

Algorithm to optimize parameters of a model, by trying at random within predefined ranges. During the search the ranges are shrunk to speed up the search process.

Cross validation (all-but-one method)

An efficient way of testing a model is to predict each lake in the database, using all other lakes.

Dissimilarity

The dissimilarity between the question lake and all other lakes is calculated by a dissimilarity index. The Euclidean Distance (ED) is the most frequently used index. It is the distance in the n-dimensional space constrained by the conditional variables (Each variable is one dimension of the space).

Distance weighting power

Data points that are used to make a prediction of the response variable of the question lake are weighted such that the influence of the data points declines with the dissimilarity from the point being estimated. The more negative the weighting power, the faster the decline in influence and the less the effect of points further out will have on the interpolation.

Input screen

Screen to select conditional and response variables.

LABDA

Largely Analogous But Differences Also, an approach of predicting the responnse of a lake. It involves two steps: (1) crude estimation using analogies; (2) fine tuning of the prediction using models or knowlegde of processes.

Logical expressions

A logical expression consists of one or more relational expressions that are combined by use of logical operators.

Logical operators

Logical operators are used to combine two or more relational expressions. The following operators are supported:

not	negation	and	logical and
or	logical or	xor	logical exclusive or

Nominal variables

Nominal variables are class variables. Values assigned to these classes do not imply any order. Differences between them have no function.

Number of nearest points

The number of cases on which the prediction is based. If you want to select all available points enter n=0.

Missing values

If one or more variables are not known, no dissimilarity can be calculated. These lakes are discarded in the calculated. In the data-set these values are marked with value -9999.9999.

Ordinal variables

Ordinal variables are class variables. Values assigned to the classes imply an order. Differences have no absolute meaning.

Question lake

A lake with known conditions but with an unknown response.

Protected mode

The protected mode is a mode of operation of a 80286 or later processor. This mode is not compatible with earlier processors, because it extends the addressing up to 16 Mb. Protected mode programs that are developed in Borland Pascal use a special run time manager (RTM.EXE) that manages the memory usage. A 80286 or later processor and at least 2Mb extended memory is required.

Real mode

The real mode is a mode of operation of a processor of the PC. This mode is compatible with the original 8086 processor. Only 1 Mb of memory can be addressed in this mode.

Relational expression.

A relational expression is a simple logical expression. It consist of:

1. a variable name
2. a relational operator (e.g. = > < <>)and
3. a value or class name. Class names (e.g. 'Sand') must be written between delimiters (').

Relational operators

The following operators can be used in relational expressions:

=	equal	<>	not equal
>	greater than	<	less than
>=	greater than or equal to	<=	less than or equal to

Response variables

Response variables are variables that express the state of the ecosystem of a lake. In regression these variables are called dependent variables.

Examples of response variables in limnology:

Secchi-depth	Chlorophyll
Algae composition	Fish species

Scroll bars

If not all information can be displayed, vertical scroll bars are shown. You can use this scroll bars with a mouse to scroll the contents of the screen.

Start-up file (default extension *.INI)

In this ASCII file the colors of BASIS can be changed and the names of the input and output files. Use an editor to make changes. The commands are not case sensitive (see Appendix 1).

Status line

The status line is a bar at the bottom of the screen, where information and hot keys are displayed. The hot keys can also be activated by a mouse click on the key name or the text after the key name.

Standardization

Give the same weight to different scaled variables.

Transformation (non-linear)

Transform variables to change the scale of a variable, such that certain parts of the scale are shrunk while other parts are stretched.

Surfer

SURFER is a registered trademark of Golden Software, Inc., P.O. Box 281, Golden, Colorado 80402 USA. BASIS can produce Command files to be used by SURFER and grid files in the format of SURFER. If SURFER is present on your PC and there is enough free memory, three dimensional Surfer plots can be made automatically (Option WAP relations|3D WAP relation). These plots can be rotated on your screen, without leaving BASIS. Otherwise the grid files can be used by other programs.

Values of the question lake

The values of conditional variables of the question lake have to be entered, if a prediction of the response variables is to be made.

WAP relations

The prediction method of the program can also be used as a kind of running avegage method of displaying the relation between two or more variables. We call this WAP relations.

WAP

WAP stands for Weighted Analogies Prediction. It is a technique to find analogies and to predict the response of an unknown case.

Weights

Less important conditional variables can be down-weighted by the user by giving them a lower weight then the other variables.

By default all weights have the value 1.

The weights can also be optimized automatically by a parameter optimization technique (Controlled Random Search).

5. REFERENCES

de Hoop, B.J., P.M.J. Herman, H. Scholten, K. Soutaert, 1992. Seneca 1.5. A Simulation ENvironment for ECological Application. Manual. Netherlands Institute of Ecology, Yerseke The Netherlands. ISBN 90-9004976-2.

Jongman, R.H., C.J.F. ter Braak & O.F.R. van Tongeren, 1987. Data analysis in community and landscape ecology. - Pudoc Wageningen.

Klepper, O. & E.M.T. Hendrix. submitted A comparison of algorithms for global characterization of confidence regions for nonlinear models. J. Env. Toxicol. Chemistry.

Sas, H. (ed.), 1989. Lake restoration by reduction of nutrient loading: expectations, experiences, extrapolations. - St. Augustin: Academia-Verl. Richarz.

Scheffer, M., 1991. On the predictability of aquatic vegetation in shallow lakes. Mem. Ist. ital. Idrobiol. 48: 207-217.

SPSS, 1989. SPSS/PC+ V2.0 Base Manual for the IBM PX/XT/AT and PS/2. SPSS Inc. Chicago.

Sokal, R.R. & F.J. Rolff, 1981. Biometry, The Principles and Practice of Statistics in Biological Research. 2nd edition. - W.H. Freeman and Company, New York.

Surfer, Surfer Reference manual.

Trimbee, A.M. & E.E. Prepas, 1987. Evaluation of total phosphorus as a predictor of the relative biomass of blue green algae with emphasis on Alberta Lakes. Can. J. Fish. Aquat. Sci. 44: 1335 - 1342.

APPENDICES

Appendix 1. Start-up file

Start-up file (default extension *.INI)

In this ASCII file the colors of BASIS can be changed and the names of the input and output files. Use an editor to make changes. All commands are not case sensitive.

The following commands are supported:

*	Comment line
DefColor	Normal text color
Monochrome	Monochrome mode (True or False)
ShadowColor	Color of shadows
HighColor	Highlighted color
DefBkColor	Background color
DefFillColor	Color inside box of screen
LogoColor	Color of the logo
LogoFillColor	Background of the logo
CursorColor	Color of the cursor
WapColor	Color of wap relation
DataFile	Name of binary file with data
DbfVarsFile	Name of dBase file with variables
DbfClassFile	Name of dBase file with data
DbfDataFile	Name of dBase file with data
OutFile	Name of output text file
WriteOutput	Output written to ASCII file ?(Yes/No) (Optimization report, test prediction, and others can produce ASCII output)
GridFile	Name of output grid file (for surfer)
NLines	Number of lines in input screen
Condition	Any condition, e.g., Sediment='Sand' to select only lakes with sandy sediment.

Possible colors:

Black	Blue	Green
Cyan	Red	Gray (VGA only)
Magenta (not VGA)	Brown	LightGray
Darkgray	LightBlue	LightGreen
LightCyan	LightRed	LightMagenta
Yellow	White	

BASIS.INI is the default start-up file with defaults. To use a different start-up file (for example BASMONO.INI), start BASIS in DOS the following way:

BASIS BASMONO

Appendix 2. How to enter new data in the database?

BASIS uses a binary file for its data input. It is however possible is to enter the data in three dBASE III files, which will be transformed into a binary file:

1. dBASE file with variables

The structure of this file is as follows:

Field	Field name	Type	Width	Dec	Index
1	VARIABLE	Character	10		N
2	NAME	Character	35		N
3	UNIT	Character	5		N
4	VARMIN	Numeric	10	3	N
5	VARMAX	Numeric	10	3	N
6	DEFTRANSFO	Numeric	1		N

Note: the contents of the field VARIABLE must be the exact the same as the field names in the file with lake information (3).

2. dBASE file with variable labels

The values of class variables can be transformed to labels. The unit of these variables (see 1) must be 'CODE'.

The structure of this file is as follows:

Field	Field Name	Type	Width	Dec	Index
1	VARIABLE	Character	8		N
2	VALUE	Numeric	2		N
3	LABEL	Character	10		N

3. dBASE file with lake information

The structure of this file is as follows:

Field	Field name	Type	Width	Dec	Index
1	LAKE	Character	8		N
2	NAME	Character	35		N
3	YEAR	Numeric	4		N
4	VAR1	Numeric	10	5	N
...					
N+3	VARN	Numeric	10	5	N

VAR1 to VARN contains the values of the N variables. Beware: the number of variables must be the same as in the file with the variable information and the field names should be the same as the contents of the first field in file 1.

To enter these new data change the file names in the start-up file (see: Start-up file (*.INI)) and start BASIS. If the new binary file is not found it automatically creates this file by reading the dBASE files.

Appendix 3. Format of the grid files

SURFER is a registered trademark of Golden Software, Inc., P.O. Box 281, Golden, Colorado 80402 USA.

BASIS can produce Command files to be used by SURFER and grid files in the format of SURFER. If SURFER is present on your PC and there is enough free memory, three dimensional Surfer plots can be made automatically (Option WAP relations|3D WAP relation). These plots can be rotated on your screen, without leaving BASIS. Otherwise the grid files can be used by other programs.

The format of the ASCII grid file (*.GRD) is as follows (Surfer, Reference Manual):

id id (4 characters) always: DSAA, which means ASCII grid file.

nx,ny nx = number of grid lines along X axis (columns).
ny = number of grid lines along Y axis (rows).

xLo,xHi xLo = minimum X coordinate of grid.
xHi = maximum X coordinate of grid.

yLo,yHi yLo = minimum Y coordinate of grid.
yHi = maximum Y coordinate of grid.

zLo,zHi zLo = minimum Z coordinate of grid.
zHi = maximum Z coordinate of grid.

grid row 1,

grid row 2

..

grid row ny Z values of the grid organized in row order. Each row has a constant Y coordinate, with the first row equal to ylo, and the last row yhi. X coordinates within each row range from xlo to xhi.

Appendix 4. Error messages

Error 1. Out of memory.

Not enough memory to run BASIS.

Error 2. A nominal variable may not be selected as response variable.

Select another variable.

Error 3. Not a valid value entered.

Acceptable numbers are:

123.232

12E3

-1223.2

Error 4. Negative weight.

The weight of a variable should be a positive value, otherwise analogies are down weighted.

Error 5. Value of question lake too small.

The value of the question lake is out of the range of valid input values. Please enter a larger value, or deselect the conditional variable.

Check the units!

Error 6. Value of question lake too large.

The value of the question lake is out of the range of valid input values. Please enter a smaller value, or deselect the conditional variable.

Check the units!

Error 7. No response variables selected.

None of the variables is selected as response variable. Select the variables that should be predicted.

Error 8. No conditional variables selected

None of the variables is selected as conditional variable. Select the variables that should be used to make a prediction.

Error 9. No value entered of conditional variable

No value of one of the conditional variables is entered, therefore no predictions can be made.

Please enter a value or deselect the conditional variable

Error 10. Not enough lakes selected.

Make a new selection (Select if). Or select another binary file (Start-up file)

Error 11. Not enough lakes with valid conditional variables.

Deselect some conditional variables.

Error 13. Distance weighting power should be negative

Enter negative value

Error 14. Number of nearest points should be larger than 0.

Enter positive value

Error 15. Error in conditional expression

Many possible errors can be detected in a logical expression.

[XXXX] not found

Name of variable not found.

Class '[XXXXX]' not found

Name of a class not found

Error in number

Not a valid number

Not enough operators

More operands than operators

No relational operator

Two operands, but no relational operator

Not enough logical operators

Not enough logical operators

Too many logical operators

More logical operators than operands

[XXX] right parentheses missing

Not enough right parentheses

[XXX] left parentheses missing

Not enough left parentheses

Not enough parentheses

All logical expressions must be separated by parentheses

no message

General error in expression

Error 16. SURFER not found.

If SURFER is not available then it is not possible to get 3D-plots at run time. If you continue, an ASCII grid file is made and might be used in another GIS system (after necessary transformations).

Error 17. File not found

Enter name of an existing file or press [Esc] to cancel.

Error 18. Wrong type of File

Only files that are saved in BASIS can be loaded.

Reenter name of an existing file or press [Esc] to cancel.

Appendix 5. The database

The database consists of a selection of data per lake. From each lake one representative year is chosen to ensure statistical independency of the data. An exception on this rule is made for lakes that are in a transient phase because they were manipulated recently (mostly biomanipulation). Although different years are not fully independent, the year to year differences are expected to be large.

From some variables two means are used:

1. **May:** 'pre young-of-the-year', young planktivorous fish not yet present. Either clear water phase or a spring algal bloom.
2. **July-August:** 'post young-of-the-year', young planktivorous fish mostly present and high predation on zooplankton.

Totally there are 145 Dutch lake-years (112 lakes) in the database.
The database includes physical and chemical parameters, information on morphology, abundances of main groups of algae, zooplankton and fish and coverage with macrophytes (see Table).

In the following table all variables are listed:

VARIABLE	NAME	UNIT	DUTCH NAME
LAKE	Code of lake	-	code van het meer
NAME	Lake	-	Naam van het meer
YEAR	Year	-	Jaar
PROTECT	Protected data? (Y/N)		Beschermde gegevens (Yes/No)
MANAGER	Institute where the data came from	nomi	Instituut waar de data vandaan komen
MANIPULA	Recently manipulated ? (Yes/No)	ordi	Recent ingegrepen? (Yes/No)
YEARNO	Number of yrs after manipulation	yr	Aantal jaren na de ingreep
AREA	Area	ha	Oppervlak
MAXDEPTH	Maximum depth	m	Maximum diepte
MEANDEPT	Mean depth	m	Gemiddelde diepte
RETENTIO	Retention time class	ordi	Verblijftijd
SEDIMENT	Sediment type	nomi	Soort sediment
XCOORD	X-coordinate	km	Parijse x-coordinaat
YCOORD	Y-coordinate	km	Parijse y-coordinaat
CHLA1	Chlorophyll-a (May)	µg/l	Chlorofyl-a (mei)
CHLA2	Chlorophyll-a (Jul, Aug)	µg/l	Chlorofyl-a (juli-aug.)
CL1	Chloride (May)	mg/l	Chloride (mei)
CL2	Chloride (Jul, Aug)	mg/l	Chloride (juli-aug.)
CONDUC1	Conductivity (May)	µS/cm	EGV (mei)
CONDUC2	Conductivity (Jul, Aug)	µS/cm	EGV (juli-aug.)
NTOT1	Total-N (May)	mg/l	Totaal-stikstof (mei)
NTOT2	Total-N (Jul, Aug)	mg/l	Totaal-stikstof (juli-aug.)
PTOT1	Total-P (May)	mg/l	Totaal-fosfaat (mei)
PTOT2	Total-P (Jul, Aug)	mg/l	Totaal-fosfaat (juli-aug.)
SECCHI1	Secchi depth (May)	m	Doorzicht (mei)

VARIABLE	NAME	UNIT	DUTCH NAME
SECCHI2	Secchi depth (Jul, Aug)	m	Doorzicht (juli-aug.)
SUSPMAT1	Suspended matter (May)	mg/l	Totaal zwevende stof (mei)
SUSPMAT2	Suspended matter (Jul, Aug)	mg/l	Totaal zwevende stof (juli-aug.)
B_CYAN1	% blue-greens (Bio) (May)	%	% blauwalgen (gewicht) (mei)
B_CYAN2	% blue-greens (Bio) (Jul, Aug)	%	% blauwalgen (gewicht) (juli-aug.)
B_DIAT1	% diatoms (Bio) (May)	%	% diatomeën (gewicht) (mei)
B_DIAT2	% diatoms (Bio) (Jul, Aug)	%	% diatomeën (gewicht) (juli-aug.)
B_GREEN1	% greens (Bio) (May)	%	% groenalgen (gewicht) (mei)
B_GREEN2	% greens (Bio) (Jul, Aug)	%	% groenalgen (gewicht) (juli-aug.)
CYANNUM1	Cyanobacteria (May)	No/m	Blauwalgen (aantal) (mei)
CYANNUM2	Cyanobacteria (Jul, Aug)	No/m	Blauwalgen (aantal) (juli-aug.)
N_CYAN1	% blue-greens (No) (May)	%	% blauwalgen (aantal) (mei)
N_CYAN2	% blue-greens (No) (Jul, Aug)	%	% blauwalgen (aantal) (juli-aug.)
N_DIAT1	% diatoms (No) (May)	%	% diatomeën (aantal) (mei)
N_DIAT2	% diatoms (No) (Jul, Aug)	%	% diatomeën (aantal) (juli-aug.)
N_GREEN1	% greens (No) (May)	%	% groenalgen (aantal) (mei)
N_GREEN2	% greens (No) (Jul, Aug)	%	% groenalgen (aantal) (juli-aug.)
EMERGENT	Coverage emergent plants	%	Bedekking emergente waterplanten
SUBMERG	Coverage submerged macrophytes	%	Bedekking ondergedoken waterpl.
CLADOC1	Cladocerans (May)	No/l	Cladoceren (mei)
CLADOC2	Cladocerans (Jul, Aug)	No/l	Cladoceren (juli-aug.)
DAPHCUC1	Daphnia cuculata (May)	No/l	Daphnia cucculata (mei)
DAPHCUC2	Daphnia cuculata (Jul, Aug)	No/l	Daphnia cucculata (juli-aug.)
DAPHNIA1	Daphnia sp (May)	No/l	Daphnia sp. (mei)
DAPHNIA2	Daphnia sp (Jul, Aug)	No/l	Daphnia sp. (juli-aug.)
B_BREAM	Bream and White Bream (% kg)	%	Brasem en kolblei (gewicht)
B_CARP	Carp (% kg)	%	Karper (gewicht)

VARIABLE	NAME	UNIT	DUTCH NAME
B_COREGO	Coregonus (% kg)	%	Coregonus (gewicht)
B_OTHERS	Other fish (% kg)	%	Overige (gewicht)
B_PERCH	Perch (% kg)	%	Baars (gewicht)
B_PIKE	Pike (% kg)	%	Snoek (gewicht)
B_PIKEPE	Pikeperch (% kg)	%	Snoekbaars (gewicht)
B_ROACH	Roach (% kg)	%	Blankvoorn (gewicht)
B_RUDD	Rudd (% kg)	%	Ruisvoorn (gewicht)
B_RUFFE	Ruffe (% kg)	%	Pos (gewicht)
B_SMELT	Smelt (% kg)	%	Spiering (gewicht)
B_TENCH	Tench and Gibel Carp (% kg)	%	Zeelt en kroeskarper (gewicht)
N_BREAM	Bream and White Bream (% No)	%	Brasem en kolblei (aantal)
N_CARP	Carp (% No)	%	Karper (aantal)
N_COREGO	Coregonus (% No)	%	Coregonus (aantal)
N_OTHERS	Other fish (% No)	%	Overige vissen (aantal)
N_PERCH	Perch (% No)	%	Baars (aantal)
N_PIKE	Pike (% No)	%	Snoek (aantal)
N_PIKEPE	Pikeperch (% No)	%	Snoekbaars (aantal)
N_ROACH	Roach (% No)	%	Blankvoorn (aantal)
N_RUDD	Rudd (% No)	%	Ruisvoorn (aantal)
N_RUFFE	Ruffe (% No)	%	Pos (aantal)
N_SMELT	Smelt (% No)	%	Spiering (aantal)
N_TENCH	Tench and Gibel Carp (% No)	%	Zeelt en kroeskarper (aantal)

For the nominal and ordinal variables the following classes were defined:

VARIABLE	VALUE	CLASS LABEL (+ comments)
RETENTIO	1	0 - 0.06 year
	2	0.07 - 0.25 year
	3	0.25 - 0.50 year
	4	> 0.50 year
MANAGER	1	DK (= Denmark, currently no data included)
	2	HEW (= Zuiveringsschap Hollandse Eilanden en Waarden)
	3	HHD (= Hoogheemraadschap van Delftland)
	4	HHR (= Hoogheemraadschap van Rijnland)
	5	HHS (= Hoogheemraadschap van Schieland)
	6	HHUS (= Hoogheemraadschap van Uitwaterende Sluizen)
	7	PRF (= Provincie Friesland)
	8	PRU (= Provincie Utrecht)
	9	RIVM
	10	RWS (= Rijkswaterstaat)
	11	WZL (= Waterschap/Zuiveringsschap Limburg)
	12	ZAG (= Zuiveringsschap Amstel en Gooiland)
	13	ZSDR (= Zuiveringsschap Drente)
	14	ZWO (= Zuiveringsschap West-Overijssel)
MANIPULA	1	Yes
	2	No
SEDIMENT	1	Sand
	2	Clay
	3	Peat